Comparative metabolomics of fruits and leaves in a hyperdiverse lineage suggests fruits are a key incubator of phytochemical diversification

Gerald F. Schneider[1,2*], Diego Salazar[3], Sherry B. Hildreth[1,4], Richard F. Helm[4], Susan R. Whitehead[1]

[1] Department of Biological Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

[2] Department of Biology, Utah State University, Logan, UT, USA

[3]Department of Biological Sciences, International Center for Tropical Botany, Florida International University, Miami, FL, USA

[4] Department of Biochemistry, Fralin Life Sciences Institute, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA.

*Correspondence: jerry.schneider@usu.edu, +1(818)426-5055

Word Count: 6253

- Introduction: 1077
- Materials and Methods: 2968
- Results: 1156
- Discussion: 1052

Figures: 6, all in color

Tables: 4

Supporting Information

- Figures: 1, in color
- Tables: 1

Summary

- Interactions between plants and leaf herbivores have long been implicated as the major driver of plant secondary metabolite diversity. However, other plant-animal interactions, such as those between fruits and frugivores, may also be involved in phytochemical diversification.
- Using 12 species of *Piper*, we conducted untargeted metabolomics and molecular networking with extracts of fruits and leaves. We evaluated organ-specific secondary metabolite composition and compared multiple dimensions of phytochemical diversity across organs, including richness, structural complexity, and variability across samples at multiple scales within and across species.
- Plant organ identity significantly influenced secondary metabolite composition, both independent of and in interaction with species identity. Leaves and fruit shared a majority of compounds, but fruits contained more unique compounds and had higher total estimated chemical richness. While organ-level chemical richness and structural complexity varied substantially across species, fruit diversity exceeded leaf diversity in more species than the reverse. Furthermore, the variance in chemical composition across samples was higher for fruits than leaves. By documenting a broad pattern of high phytochemical diversity in fruits relative to leaves, this study lays groundwork for incorporating fruit into a comprehensive and integrative understanding of the ecological and evolutionary factors shaping secondary metabolite composition at the whole-plant level.

Key words: secondary metabolites, chemical diversity, metabolomics, fruit, seed, leaf

Introduction

Phytochemistry plays a key role in mediating the ecological and evolutionary dynamics of plant interactions (Kessler & Baldwin, 2002; Wittstock & Gershenzon, 2002; Hartmann, 2007). As functional traits, secondary metabolites can significantly affect plant fitness by defending plants against antagonists, directly affecting the competitive ability of neighboring plants, protecting plants from harsh environmental conditions, and attracting and rewarding mutualists, both above and below ground (Iason *et al*., 2012). However, research on secondary metabolites and their role in the ecology and evolution of plants has been disproportionately focused on vegetative organs, specifically the leaf (e.g. Kursar *et al*., 2009; Richards *et al*., 2015; Volf *et al*., 2018; Salazar *et al*., 2018). While secondary metabolites have numerous demonstrated functions mediating plant-animal interactions surrounding leaves, they also likely perform a crucial and complex set of functions in reproductive organs.

Plant reproductive organs have been a nexus of plant-animal interactions since before the emergence of angiosperms. However, the ecological role that secondary metabolites play in the biology of these plant organs has not been deeply explored. Fruits, and the seeds they contain, provide a direct link to plant fitness and are therefore likely to be under intense

selection pressure to attract mutualists and deter antagonists. These complex and contrasting selective pressures are distinct from those acting on leaves, and may lead to the occurrence of secondary metabolites not found in other organs. Indeed, given the complex and often contrasting nature of selective pressures to which fruits and seeds are exposed, fruits and seeds are likely to serve as evolutionary incubators of novel secondary metabolites, and disproportionately contribute to the diversity of phytochemical traits. This is especially likely in systems involving animal-mediated seed dispersal (zoochory), in which plants face the ecological and physiological challenge of attracting and offering a nutritional reward to dispersal vectors while also repelling seed predators, pathogens, and non-target frugivores (Herrera, 1982; Tewksbury, 2002; Whitehead *et al*., 2016).

Secondary metabolites endemic to fruits, and with demonstrated functional significance in seed dispersal and/or fruit defense, have been shown in several systems, including iridoid glycosides in honeysuckles (Whitehead & Bowers, 2013a, 2013b), capsaicinoids in *Capsicum* (Suzuki & Iwai, 1984; Tewksbury & Nabhan, 2001; Tewksbury *et al.*, 2008), and amides and alkenylphenols in *Piper* (Whitehead *et al*., 2013, 2016; Whitehead & Bowers, 2014; Maynard *et al*., 2020). Further, capsaicinoids in *Capsicum* and alkenylphenols in *Piper* are synthesized only in the fruits of these taxa (Suzuki & Iwai, 1984; Maynard *et al*., 2020). Overall, these studies suggest that unique and potentially contrasting selective pressures on fruits may be an important factor shaping phytochemical diversification in plants. However, our understanding of the relative importance of interactions across plant organs in shaping phytochemical diversity is limited by a paucity of studies that compare chemical composition and metabolomic diversity across plant organs in an ecological context.

Comparative metabolomic studies across plant organs have the potential to greatly expand our understanding of secondary metabolite function and evolution. Given that metabolites may be organ-specific, the location in which they are expressed in the plant (and consequently, the ecological interactions in which they are involved) can provide valuable insight into both the evolutionary origins and ecological consequences of the vast diversity of undescribed plant secondary metabolites.

Despite the likelihood of distinct selective pressures promoting divergent evolution of secondary metabolites across plant organs, it is likely that the phytochemical diversity in one organ may be constrained by physiological or genetic linkages with the phytochemistry of other organs (Adler *et al.*, 2006, 2012; Kessler & Halitschke, 2009; Keith & Mitchell-Olds, 2019). Physiological constraints may result when a majority of the steps in a secondary

metabolite pathway are localized to a particular part of the plant, yielding complete or nearly complete end products that are then transported to the organs in which they are utilized, e.g. glucosinolates in the Brassicaceae (Keith & Mitchell-Olds, 2019). Such a pathway has a limited capacity to generate organ-specific modifications of its end products prior to transport, and the sink organs may lack the metabolic machinery required for such modifications. Other secondary metabolites are locally synthesized, but in this case organ-specific metabolites derived from a shared metabolic pathway may be limited by genetic linkage, through co-localization of genes responsible for modifications within a metabolic pathway, e.g. terpene synthase clusters (Falara *et al*., 2011; Chen *et al*., 2020; Xu *et al*., 2020). Certainly, evolutionary processes may overcome these constraints when there are conflicting selection pressures among organs, as evidenced by the numerous examples above of compounds occurring only in specific organs. Furthermore, even when fruits and leaves do share compounds, these compounds may be quantitatively uncorrelated (Cipollini *et al.*, 2004; Whitehead & Bowers, 2013; Berardi *et al*., 2016). Thus, while all plant species are biochemically circumscribed to some extent by the biosynthetic pathways acquired through their evolutionary history, broad evolutionary patterns of such constraints across plant organs have yet to be elucidated. Comparative metabolomics provide us with the tools to define and characterize these patterns of constraint in conjunction with patterns of phytochemical innovation.

In this study, we use comparative untargeted metabolomics to explore whether and how differential selective pressures and constraints across reproductive and vegetative organs have shaped the diversity and distribution of secondary metabolites in *Piper*, a pantropical species-rich genus. *Piper* are diverse and dominant members of neotropical lowland forest understories and are known to contain a rich array of secondary metabolites (Kato & Furlan, 2007; Richards *et al.*, 2015). Their well-studied chemical composition and a long history of ecological research have made them a model system for understanding phytochemical diversification and its role in shaping plant interactions and community structure (Dyer & Palmer, 2004; Richards *et al.*, 2015; Salazar *et al.*, 2016).

Our overall objective in this study is to test the hypothesis that fruits can act as incubators of phytochemical diversification in plants. First, we describe the occurrence patterns of secondary metabolites across leaves, fruit pulp, and seeds in 12 *Piper* species, providing baseline data for understanding *Piper* secondary metabolite function. We use untargeted mass spectrometry-based metabolomics, molecular networking, and in-silico fragmentation modeling to characterize undescribed metabolites, followed by machine

learning and distance-based methods to compare composition across organs and species. Second, we use these data to test predictions of high relative diversity in fruits derived from our hypothesis of fruit-driven phytochemical diversification. We compare multiple dimensions of phytochemical diversity across leaves and fruit organs, including the richness at multiple scales (alpha and gamma diversity), variability (beta diversity), and structural complexity of secondary metabolites.

Materials and Methods

Study system

Encompassing over 1,000 species across the Neotropics (Quijano-Abril *et al.*, 2006), the genus *Piper* is diverse and abundant in forest understories, clearings, and edges (Gentry, 1990; Dyer & Palmer, 2004). *Piper* growth forms range from herbs and vines to shrubs and small trees (Gentry, 1990; Dyer & Palmer, 2004). Fruits of Neotropical *Piper* are borne on distinct spike-shaped infructescences that are dispersed primarily by bats of the genus *Carollia* (Phyllostomidae). Fruit antagonists of *Piper* include insect seed predators, which have been found to consume up to 87% of seeds (Greig, 1993), and a largely uncharacterized suite of pathogens, which rapidly attack fruit upon ripening (Thies & Kalko, 2004; Whitehead & Bowers, 2014; Maynard *et al*., 2020). Leaves of *Piper* are subject to herbivory from a broad array of arthropods, including a genus of specialist geometrid moths, *Eois*, estimated to include over 1,000 species in the Neotropics (Brehm *et al.*, 2016), as well as other geometrid moths, coleopterans, and orthopterans (Dyer & Palmer, 2004).

Field collections

All field collections took place between 2009 and 2012 at La Selva Biological Station, Heredia Province, Costa Rica. Samples were collected during a phenology census across 28 species of *Piper* during 2009-10 and opportunistically from 2010-12 when ripe fruits were available. Ripe fruits were distinguished by a distinct softening and swelling of the fruit along an infructescence combined with a partial senescence of the infructescence from the branch (presumably to allow bats to easily remove the entire infructescence in flight). In most *Piper* species included in this study, one or a few infructescences ripen per day per plant during the fruiting period, and the vast majority of these are removed on the same night of ripening by bats (Thies & Kalko, 2004; Maynard *et al*., 2020). Those that are not removed rapidly decompose; therefore, we always took care to collect freshly-ripened infructescences. We chose 12 species for inclusion in this study for which we were able to obtain collections from at least three individual plants. For each individual, we collected 1-2 ripe infructescences and the unripe infructescences that were immediately distal to the ripe ones on the same branch.

Fruits on a *Piper* branch mature sequentially from the proximal to the distal end of the branch; thus, these adjacent unripe infructescences were the next closest to maturity on that branch. Leaves were collected from the same branch. We chose the youngest fully-expanded leaf that did not have extensive herbivore damage. All samples were transported immediately to the laboratory (within 2 hours) and frozen at -80ºC prior to analysis. Subsequent analyses involve four sample types: complete leaves, pulp from unripe and ripe infructescences, and seeds from ripe infructescences.

Chemical extractions

The frozen plant material was freeze-dried (-20° C/ -55 ° C, shelf/condenser), then ground to a fine powder using a FastPrep-24 homogenizer. Seeds and pericarp were separated prior to grinding by gently rubbing the dried fruit over fine mesh; the lignified central rachis of the infructescence was discarded. In unripe fruit, seeds that were not sufficiently developed to be separated from the pericarp by this method were homogenized with the pericarp. For each sample, 50 mg of homogenized powder was weighed into a 2 mL Eppendorf tube using a microbalance. To isolate the broadest possible range of phytochemicals while excluding the broadest possible range of primary metabolites, extracts were prepared using buffered acetonitrile and acetone in series. The acetonitrile and acetone extraction solutions were prepared with an aqueous acetate buffer (44.3 mmol/L ammonium acetate), both at 70:30 solvent: buffer, v/v. The solutions were prepared with Nanopure® water, Fisher HPLC-grade acetic acid, and Fisher Optima®-grade ammonium acetate, acetonitrile, and acetone. All containers and instruments coming into contact with the extracts were rinsed with Fisher Optima®-grade methanol. Each 50 mg sample was extracted twice with 1.5 mL buffered acetonitrile, then twice more with 1.5 mL buffered acetone (6.0 mL total extraction solution). During each of these four extractions, the sample was mixed with the extraction solvent for 5 min in a vortexer, and then centrifuged for 5 min at 15870 rcf, after which the supernatant was removed and added to a 20 mL glass scintillation vial. The supernatant from each of the four extractions was combined in the same 20 mL vial. The combined extract was dried at 30° C using a nitrogen evaporator until no solvent was visible, then further dried in a lyophilizer for 12 h (-20° C/ -55 ° C, shelf/condenser) before being transferred to storage at -80° C until analysis.

Untargeted metabolomics

LC-MS data were collected using an Acquity I-class UPLC coupled to a Waters Synapt G2-S quadrupole time-of-flight mass spectrometer (Waters). For analysis, dried extracts were resuspended at 10 mg/mL in 75:25 water: acetonitrile + 0.1 % formic acid, with 1.0 µg/mL

N-oleoylglycine as an internal standard. The extract was then sonicated for 10 min, after which a 20 µL aliquot was taken and diluted 10-fold with 75:25 water: acetonitrile + 0.1 % formic acid. The diluted aliquot was then vortexed and centrifuged (10 min,13,000 xg) and an aliquot (180 µL) was transferred to an LC-MS vial for analysis. Solvent blanks and combined, quality-control samples were injected at regular intervals during data collection. The autosampler temperature was 10°C and the injection volume was 1.5 µL. The column employed was a reverse-phase Acquity BEH C18 (2.1 mm ID x 150 mm, 1.7 um particle size, Waters) maintained at 35 °C at a flow rate of 0.2 mL/min. Solvent A was water with 0.1% formic acid and solvent B was acetonitrile with 0.1% formic acid (LCMS grade, Fisher Chemical). Solvent gradient: 0-0.5 min, 90% A; 0.5-1.0 min, 75% A; 1.0-8.0 min, 5% A; 8.0-10.0 min, held at 5% A; 10.0-11.0 min, 90% A; 11.0-15.0 min, held at 90% A. Mass spectra and fragmentation spectra were collected simultaneously using Waters' MS^E in positive-ion mode, with the following parameters: peak data recorded in centroid mode; 0.185 s MS scan time; 20-35 V collision energy ramp; argon collision gas; 125° C source temperature; 3 V capillary voltage; 30 V sample cone voltage; 350° C desolvation temperature; nitrogen desolvation at 500 L/hr; 10 µL/min lockspray flow rate; 0.1 s lockspray scan time; 20 s lockspray scan frequency; 3 lockspray scans to average; 0.5 Da lockspray mass window; 3 V lockspray capillary voltage. The lockspray solution was 1 ng/ µL leucine enkephalin, and sodium formate was used to calibrate the mass spectrometer.

Alignment, deconvolution, and annotation of molecular and adduct ions were conducted using the XCMS and CAMERA packages in R statistical software (Smith *et al.*, 2006; Tautenhahn *et al.*, 2008; Benton *et al.*, 2010; Kuhl *et al.*, 2012) with parameters in (R code repository).

Molecular networking

Molecular networking was used to quantify and visualize the dimensions of the chemical structural trait space occupied by the secondary metabolites in our study (Aron *et al.,* 2020). This technique employs tandem mass spectrometry to generate fragmentation spectra for each putative compound. These fragmentation spectra are diagnostic of molecular structure, and through pairwise comparison they are used to generate a network linking putative compounds to one another based on structural similarity.

In our study, fragmentation spectra data files were aligned, deconvoluted, and converted to .mgf using MS-DIAL software (v4.10) and were then uploaded to the Global Natural Products Social Molecular Networking (GNPS) online workflow for molecular networking and library-based annotation. The following parameters were used for the GNPS

workflow METABOLOMICS-SNETS-V2 (v14): 0.02 Da precursor ion mass tolerance; 0.02 Da fragment ion mass tolerance; minimum matched fragment peaks = 6; minimum cluster size = 3; minimum cosine score for network pairs = 0.7; network TopK = 1000; maximum connected component size = 0. All mass spectral libraries available through GNPS which contained data collected in positive ion mode were used for annotation. Library search parameters were: minimum matched peaks = 6; cosine score threshold = 0.6; maximum analog mass difference = 100. Workflow options for advanced filtering, advanced GNPS repository search, and advanced output were not used.

For further annotation via *in-silico* modeling, results of the METABOLOMICS-SNETS-V2 workflow were passed to a second GNPS workflow, Network Annotation Propagation (NAP_CCMS v1.2.5). The parameters used for NAP_CCMS were as follows: all clusters selected; subselection cosine value = 0.7; first candidates for consensus score = 10; fusion results used for consensus; accuracy for exact mass candidate search = 15 ppm; acquisition mode = positive; adduct ion types = $[M+H]^+$ and $[M+Na]^+$; all structure databases selected; no custom database or parameter file; compound class not specified; parent mass selection enabled; maximum number of graphed candidate structures = 10; standard workflow type.

Finally, the outputs from METABOLOMICS-SNETS-V2 and NAP_CCMS were combined and exported for visualization using the GNPS workflow MolNetEnhancer (v15). Network visualization and curation was conducted using Cytoscape software (v3.7.2). Parent masses of features in the molecular network were curated based on the XCMS-CAMERA output described above, with primary metabolites and artefactual or pseudoreplicated features removed from the network and subsequent analyses. Features in the molecular network were annotated to the level of chemical class, e.g. flavonoid or prenol lipid, based on ClassyFire chemical taxonomy as applied by MolNetEnhancer. The list of annotated molecular features returned by XCMS-CAMERA processing was used to compare overall phytochemical composition across organs and species.

Unfragmented ions collected during single-mass-spectrometry and subsequently aligned, deconvoluted, and annotated, as described above, were used to compare overall phytochemical composition across organs and species. Ion abundance data were transformed to presence/absence data using the peak recognition parameters in XCMS (R code repository). Ion presence/absence was used for analyses rather than relative ion abundance for two reasons: 1) our sample size affords limited capacity to account for variation in abundance within a given organ of a given species, and 2) the scale of variation in ion

abundance is likely to differ widely across the structurally diverse compounds in *Piper* due to variation in ionization efficiency (Cech & Enke, 2001).

Comparisons of phytochemical composition across organs and species

To compare metabolome-level patterns of phytochemical composition across organs and species, we conducted two separate analyses of the multivariate sample composition, focused first on compound occurrences (presence/absence data) and second on the structural composition of samples. We focused on the occurrence and structure of each molecular feature and omitted information on relative abundances due to the infeasibility of accounting for variation in ionization efficiency across hundreds of uncharacterized compounds. First, to visualize differences in patterns of compound occurrence across samples, we used non-metric multidimensional scaling (NMDS) based on the Sørensen dissimilarity index (binary Bray-Curtis). We then tested for effects of organ, species, and their interaction on compound composition using PERMANOVA, implemented with the 'adonis2' function in the R package 'vegan'. The individual plant identity was included in these analyses as a 'strata' (i.e. random effect), and we used 999 permutations (note that this means the minimum possible *P*-value is $P = 0.001$, indicating that the observed differences in sample composition could not be replicated in any of the 999 permutations). To further understand specific differences among the four organ types, we followed this analysis with post-hoc pairwise PERMANOVAs for all possible combinations of organ types, correcting for multiple comparisons using the 'pairwise.adonis2' function (Martinez Arbizu, 2020). In addition, based on strongly supported interactions between organ and species (see results), we also divided the data by species and tested for the effects of organ on compound composition for each species individually. All analyses were conducted using the 'vegan' package in R (Oksanen *et al.*, 2019).

In addition to our analysis of compound occurrence, we also examined how the structural composition of samples was affected by organ, species, and their interaction. To account for structural features, we generated a multivariate structural dissimilarity index that was a modification of Sedio *et al.*'s (2017) Chemical Structural and Compositional Similarity (CSCS) index, which quantifies the pairwise similarity of samples by calculating the maximum cosine similarity of the aligned MS-MS ion fragmentation spectra for each inter-sample pair of molecular features. We modified this index by representing ion abundance as a binary term and expressing the index in terms of dissimilarity (1-CSCS). The structural dissimilarity matrix was then used as the basis for NMDS and PERMANOVAs as above that examined the effects of organ, species, and their interaction on structural composition.

Machine learning

To identify molecular features that distinguished different organs, we used random forest analysis via the "randomForest" and "Boruta" packages for R statistical software (Liaw & Wiener, 2002; Kursa & Rudnicki, 2010). All molecular features distinguished in XCMS-CAMERA processing were used as variables in these analyses. The random forest analysis used a decision tree model to assign samples to our four organ groups (Breiman, 2001; 2002). In the process, the analysis ranked molecular feature variables according to their importance in the model's group assignments. Boruta analysis complemented the bottom-up random forest analysis by applying a top-down search for molecular features that were important in informing group assignments. This is accomplished by comparing the features' importance with importance achievable at random, using "shadow" variables which are generated by permuting the original variables (Kursa & Rudnicki, 2010).

Comparisons of chemical diversity across organs

Phytochemical diversity is a multifarious concept that includes the number of compounds (richness), their relative abundances (evenness), their structural complexity, and their variation in space and time (Wetzel & Whitehead, 2020). Considering the challenges associated with estimating abundances in untargeted LC-MS-MS data, we focus here on richness and structural complexity, both of which were examined at multiple scales within and across species. For each organ type, we define gamma diversity as the total diversity observed across all samples, alpha diversity as the average diversity within a single sample from one organ from one *Piper* individual, and beta-diversity as the variation (both intra- and inter-specific) across samples.

Gamma diversity. To compare the gamma diversity (total number of compounds detected across all species) of different organs, we used a rarefaction analysis analogous to those commonly used to assess species diversity (Gotelli & Colwell, 2011) with compounds as "species" as in Wetzel & Whitehead (2020). This allowed us to: 1) explicitly visualize the relationship between chemical diversity and sampling scale across different organs (i.e. alpha, beta, and gamma diversity), and 2) estimate the total compound richness in each organ type. Because our individual samples were not independent (we collected three samples per species for 12 species), we used a constrained rarefaction that is similar conceptually to spatially-constrained rarefaction (Chiarucci *et al.,* 2009). Briefly, samples were added to bootstrapped accumulation curves in a semi-random manner in which samples from the same species were grouped. For each iteration, a random sample was chosen as a starting point, then other samples from that species were added in random order prior to choosing another sample at

342 random, following with all other samples from that species, and so on until all species were
343 included. We estimated total species richness from these curves using the 'fitspecaccum'
344 function in 'vegan' based on an asymptotic regression model. Accumulation curves and fits
345 were averaged across 5000 bootstrapped samples with random starting points.
346 *Alpha diversity.* To compare the average compound richness in a sample (i.e. alpha diversity)
347 across organs, we used a linear mixed model with organ, species, and their interaction as
348 fixed effects and plant identity as a random effect. For hypothesis testing, we compared the
349 full model to simplified versions with fixed effects terms deleted using likelihood ratio tests.
350 Based on a strong interaction between organ and species (see results), we further divided the
351 data by species and examined differences in richness among organs for each species
352 separately.
353 *Structural complexity.* To compare structural complexity across organ types, we first
354 calculated an index of structural complexity for each sample that was modified from the
355 CSCS index described in Sedio *et al*. (2017) to include only presence/absence data. This
356 within-sample CSCS represents the mean pairwise similarity among all individual molecular
357 features detected in a sample. We used the inverse of this similarity index (1-CSCS) as a
358 measure of overall structural complexity present in a sample. To examine how structural
359 complexity varied across organs and species, we used a linear mixed model with species,
360 organ, and their interaction as fixed effects and plant identity as a random effect. Hypothesis
361 testing was conducted as described above using likelihood ratio tests. Based on strong
362 interactions between organ and species (see results), we examined differences among organs
363 separately for each *Piper* species.
364 *Beta diversity.* We examined differences in beta-diversity (i.e. sample-to-sample variance in
365 composition) across organs in two ways, focusing first on variation in compound occurrences
366 (presence/absence) and second on structural features. These analyses were based on the same
367 distance matrices described above that we used to assess overall differences in composition
368 across samples, but instead focused on variance (i.e. dispersion) among samples. This was
369 assessed using the function 'betadisper' in the R package 'vegan' to compare the dispersion
370 around the group centroid across the four organ types. The 'betadisper' function calculated
371 the distances from each sample to the group centroid, and statistical support for differences in
372 dispersion across organs was assessed using a permutation test (N = 999 permutations)
373 followed by a post-hoc Tukey HSD test to assess pairwise differences among individual
374 organs. Because this analysis focused on sample-to-sample variance and our dataset included
375 multi-level sampling (multiple species and multiple individuals within species), significant

differences in beta diversity across organs could be due to both intraspecific and interspecific variance among samples. Thus, we followed this analysis with a set of PERMANOVAs, conducted separately for each organ type, with *Piper* species as an explanatory factor. This analysis allowed us to test if *Piper* species explained a significant portion of the variation in composition within an organ, and partitioned sample-to-sample variance within an organ type according to the percent of variance explained by species and the percent explained by differences among individuals within species (i.e. the residual variance).

Results

Untargeted metabolomics and molecular networking reveal high chemical diversity and many compounds unique to fruits

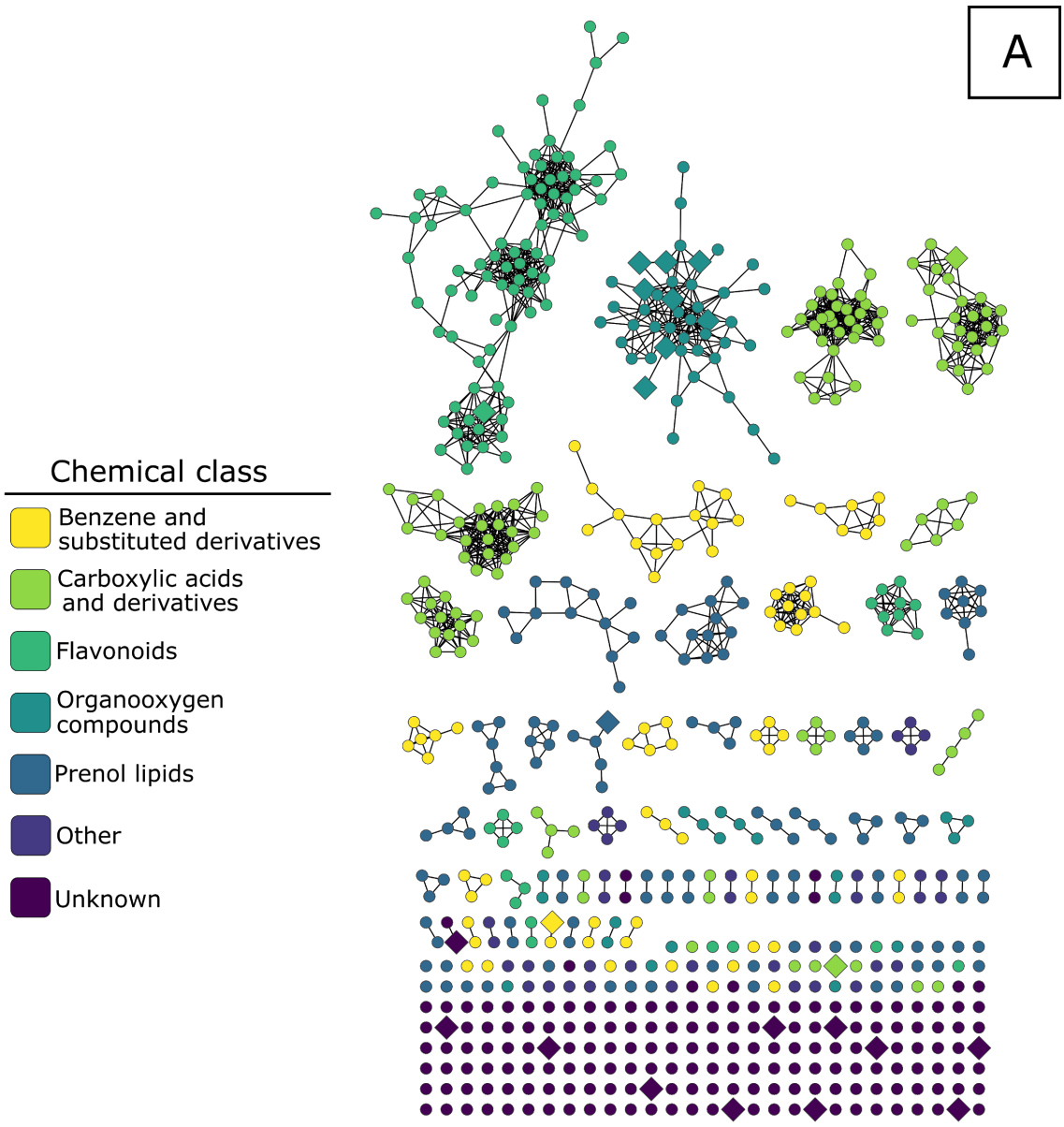
Alignment, deconvolution, and annotation of molecular and adduct ions via XCMS and CAMERA yielded 1,311 unique molecular features across all species and organs. It is important to note that, like all other metabolomic approaches, our analytical approach is likely to overestimate the true number of individual chemical compounds present in our samples. The combination of XCMS-CAMERA followed by manual curation unfortunately cannot condense all features (m/z and retention time pairs) into individual compounds. In-source fragmentation, ion clusters, centroid peak splitting of highly abundant ions, and centroid merging of ions near the noise level can all contribute to expanding the dataset beyond individual compounds. The 1,311 features described in this work thus overestimates the number of individual molecular species, though to a lesser extent than in uncurated datasets. Nevertheless, this overestimation is likely to represent a small fraction of the total chemical diversity captured in our analysis. Furthermore, this overestimation is also likely to be of equal magnitude across all species and organs and therefore, will not have a significant impact on the general conclusions of our study. Regarding terminology, these 1,311 features meet or exceed the level of curation beyond which features have, for clarity, been described as "compounds" in the chemical ecology literature (e.g.: Sedio *et al.*, 2017; Christian *et al.*, 2020; Ricigliano *et al.*, 2020). Thus, for the sake of consistency and clarity, we refer to our curated features as compounds.

Tandem mass spectrometry yielded fragmentation spectra for 706 of these compounds (Table 1, Fig. 1). Library- and *in silico*-based classification of fragmentation spectra and parent ions via GNPS resulted in annotation at the level of "class" *sensu* ClassyFire chemical taxonomy for 527 compounds in 23 classes (Table 1, Fig. 1).

Table 1: Summary of GNPS molecular network annotations. Compound richness indicates the number of putative compounds, for which fragmentation spectra were obtained, that fall under the given category. Chemical classes are per ClassyFire chemical taxonomy. A compound was labeled as fruit or leaf-specific if it was detected only in that organ within the 12 focal *Piper* species. Asterisks indicate organ-specific compound richness exhibiting $P < 0.05$ (binomial test with probability = 0.5 of occurrence in fruit or leaf; n = number of organ-specific compounds in chemical class).

Chemical Class	**Examples of class known from *Piper* spp.**	**Total Compound Richness**	**Fruit-specific Compound Richness**	**Leaf-specific Compound Richness**
Benzene and substituted derivatives	Cyanogenic benzoates, Non-prenylated benzoic acids	75	5	0
Carboxylic acids and derivatives	Amides, Chromenes, Kavalactones	122	25*	1
Flavonoids	Flavonoids	104	3	0
Organo-oxygen compounds	Oxygenated or glycosidic derivatives of other classes	65	5	0
Other	Amides, Chalcones, Chromenes, Imides	37	4	0
Prenol lipids	Chalcones, Prenylated benzoic acids, Neryl catechol diols, Terpenes	124	6*	0
Unknown		179	14*	1
Total		706	62*	2

408



409

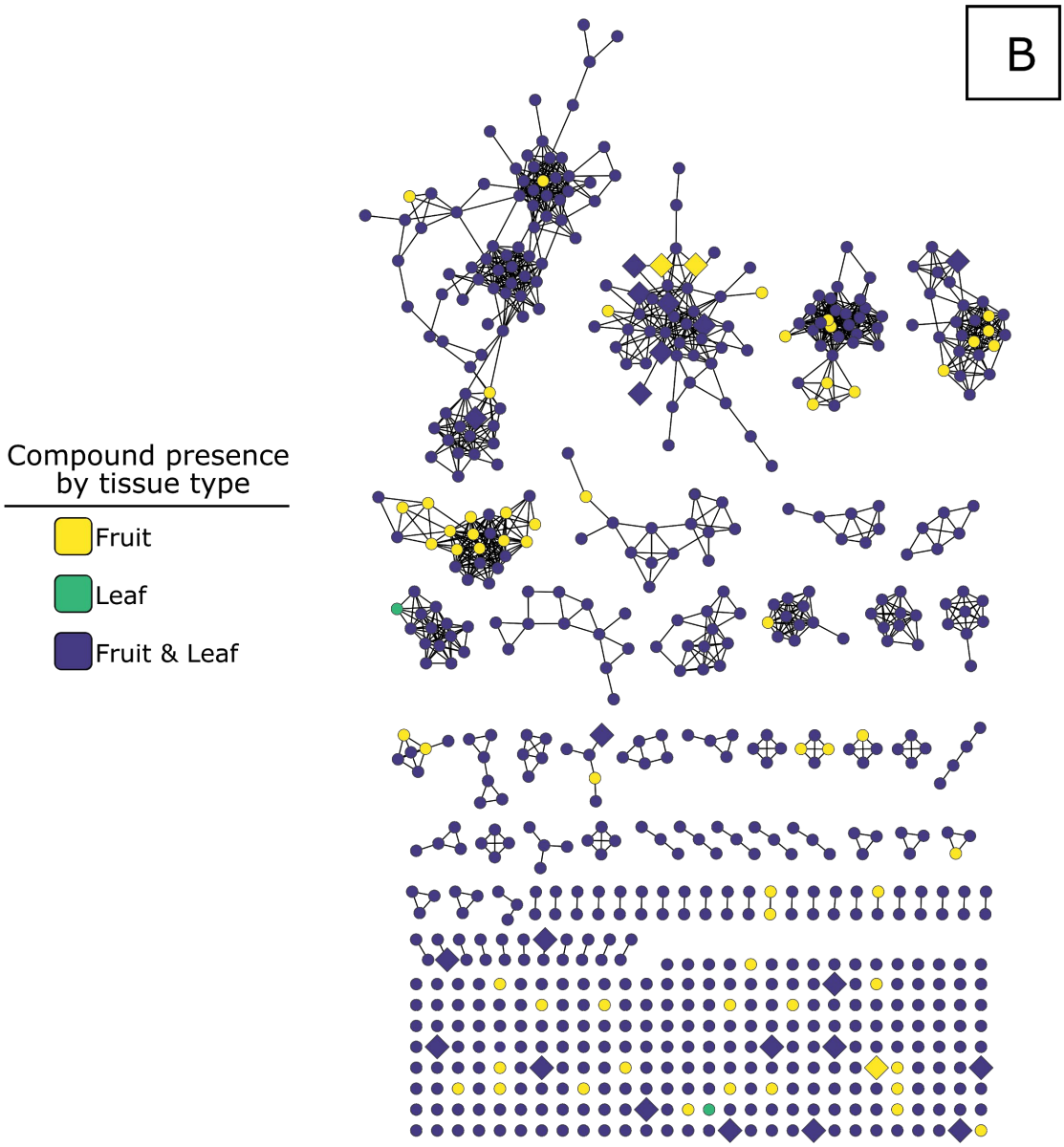


Figure 1: Molecular network of 706 compounds from 12 *Piper* species color-coded by ClassyFire chemical classification annotation (A) or by organ-level occurrence across the 12 species (B). Node and edge arrangement and compound annotation are as described in "Molecular Networking" methods. Enlarged, diamond-shaped nodes represent compounds identified by the Boruta analysis as important for distinguishing among organs. In B, compounds are coded as occurring in "fruit" if they occur in one or more of the three sample types (unripe pulp, ripe pulp, or seeds).

Phytochemical composition differs across organs and species

The multivariate patterns of phytochemical occurrence were strongly affected by organ, species, and their interaction (organ: $F_{3,95} = 24.19$, $P = 0.001$; species: $F_{11,95} = 27.65$, $P = 0.001$; organ x species: $F_{33,95} = 1.99$, $P = 0.001$; Figure 2A). Pairwise comparisons among organs indicated strong differences among all organs ($P = 0.001$ for all comparisons). Further examination of differences among organs for each of the 12 *Piper* species individually also revealed strong effects of organ in all cases (Table 2). Similarly, when we assessed factors influencing the multivariate patterns of structural composition across samples, we found a strong effect of organ, species, and their interaction (organ: $F_{3,95} = 17.34$, $P = 0.001$; species: $F_{11,96} = 21.28$, $P = 0.001$; organ x species: $F_{33,96} = 2.31$, $P = 0.001$; Figure 2B), significant differences among organs in all pairwise comparisons ($P = 0.001$ for all comparisons), and differences among organs for each individual species (Table 2).

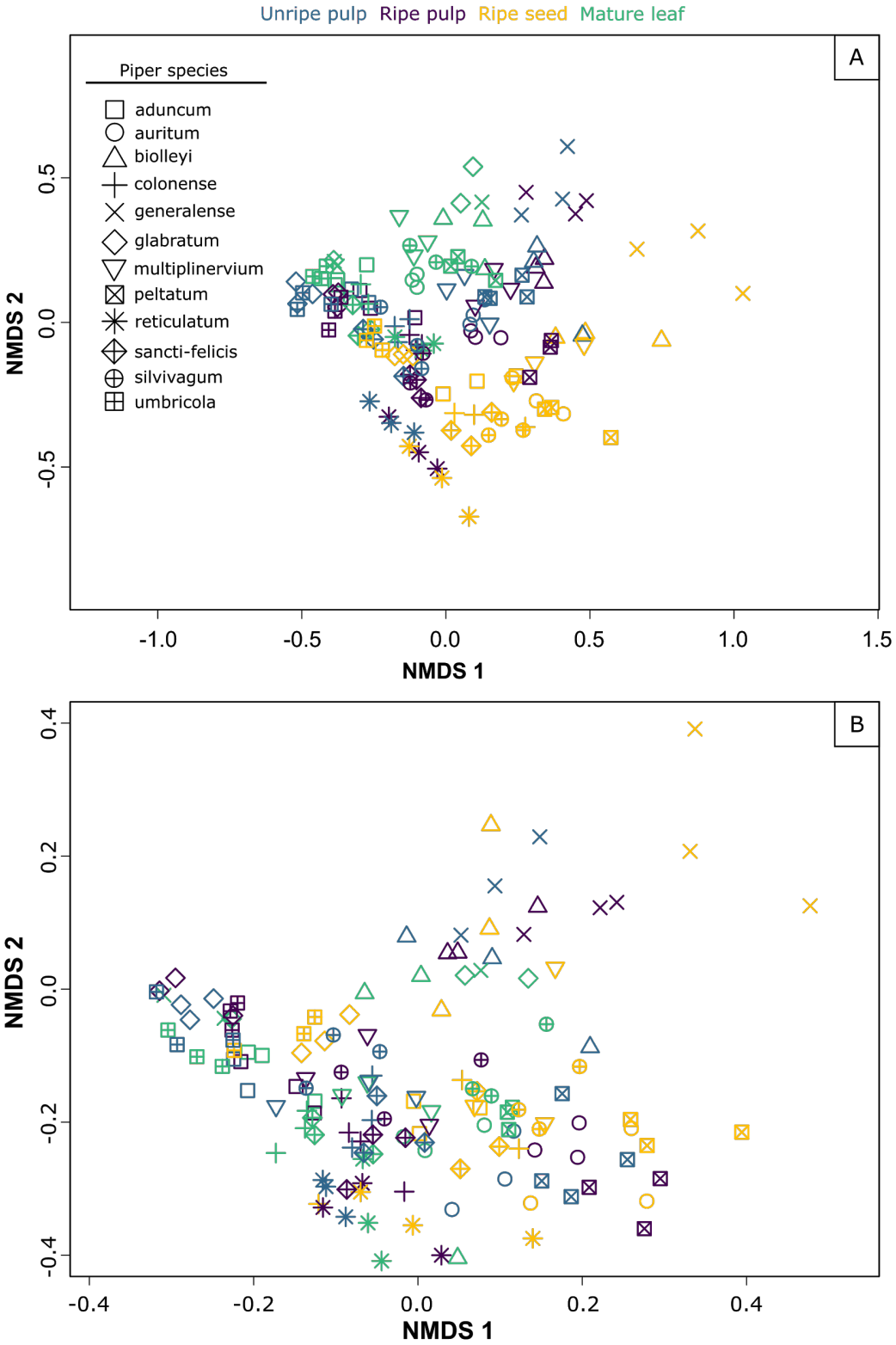


Figure 2: NMDS plots showing the effects of organ and species on two aspects of multivariate chemical composition across samples: (A) compound occurrences (presence/absence) and (B) structural composition.

439

Table 2: Results from PERMANOVAs, conducted separately for each species, testing the effects of organ type (leaves, seed, unripe pulp, or ripe pulp) on two aspects of phytochemical composition: compound occurrences and structural composition

	Compound Occurrence		**Structural Composition**	
***Piper* species**	$F_{3,11}$	***P***	$F_{3,11}$	***P***
aduncum	3.07	0.001	3.07	0.001
auritum	3.51	0.002	3.51	0.002
biolleyi	2.19	0.009	2.19	0.009
colonense	4.39	0.001	4.39	0.001
generalense	4.11	0.001	4.11	0.002
glabrescens	4.18	0.002	4.18	0.001
multiplinervum	3.45	0.001	3.45	0.001
peltatum	4.50	0.001	4.50	0.001
reticulatum	5.38	0.001	5.38	0.002
sancti-felicis	3.67	0.001	3.67	0.001
silvivagum	5.19	0.001	5.19	0.001
umbricola	2.89	0.002	2.89	0.001

440
441
442
443
444

445 ***Machine learning, informed by numerous compounds from diverse chemical classes in***
446 ***each organ, accurately distinguishes between reproductive and vegetative organs***

447 The random forest decision tree model used 2000 trees with 36 variables at each split.
448 Our analysis showed an overall out-of-bag (OOB) mean error rate of 11.72% across the four
449 organ groups. In other words, using secondary metabolites alone, the algorithm was able to
450 predict if a sample was from a leaf, ripe fruit, unripe fruit, or seed approximately 9 times out
451 of every 10 samples. Examining the error rate of each organ group, it was apparent that
452 correctly assigning pulp samples to the correct ripeness stage was the main source of OOB
453 error, with error rates of 27.78% and 18.92% for unripe and ripe pulp respectively. Leaves
454 and ripe seeds both exhibited zero OOB error. Boruta analysis, designed to both identify
455 important classification features and assess their relative contribution to the final
456 classification performance, identified 23 features exhibiting a significantly higher variable
457 importance score (VIS) than shadow variables. These 23 features are detailed in Table S1.

458 **Comparisons of chemical diversity across organs**

459 ***Gamma diversity***

460 Overall, we detected 1,311 compounds across all organ types. The large majority of
461 these compounds (1,126) were shared across all organs (Fig. 3A). Of those compounds that
462 were organ-specific, there were 92 compounds that were found only in fruits (unripe pulp,
463 ripe pulp, and/or seeds) but never in leaves, and four compounds were found only in leaves
464 but never in fruits. There were also 76 compounds that were shared between fruit pulp
465 (unripe and/or ripe) and leaves, but never detected in seeds. Rarefaction analysis showed that
466 the estimated total gamma diversity (total number of compounds across all 12 species of
467 *Piper*) was highest in unripe and ripe fruit pulp, intermediate in seeds, and lowest in leaves
468 (Fig. 3B, Table 3).

469

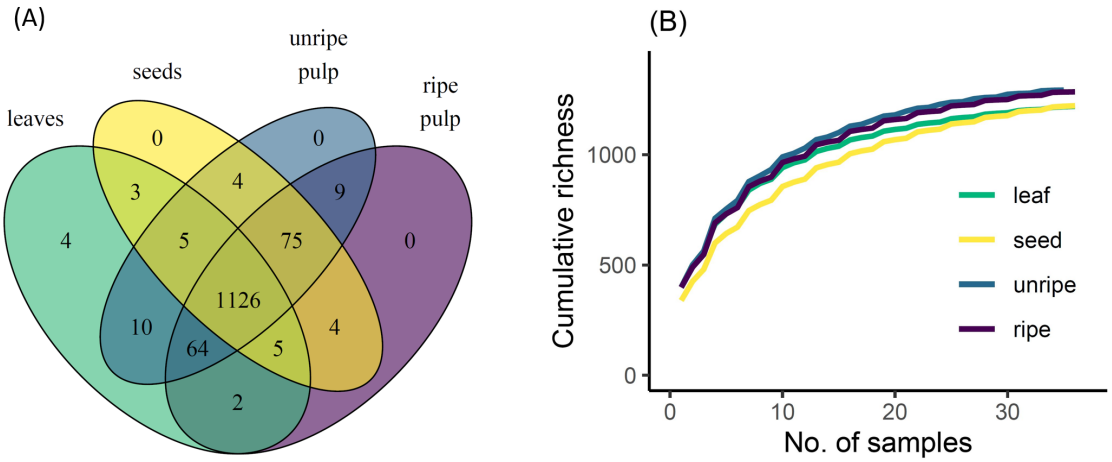


Figure 3: Chemical gamma diversity parsed by organ type. A Venn diagram (A) shows the total number of compounds detected across all samples that were unique and shared across organ type. The rarefaction curve (B) shows how compound richness accumulates with sampling scale in each organ type. Curves represent an average across 5000 bootstrapped accumulation curves with random starting points. Because samples from the same species were not independent, the rarefaction was constrained by species such that samples from the same species were always added in sequence.

Table 3: Rarefaction results showing total estimated richness for each organ across all 12 Piper species sampled

Organ	Estimated Richness	SE	95% CI high	95% CI low
leaf	1226	1.5	1229.3	1223.5
seed	1276	3.0	1282.0	1270.1
unripe pulp	1312	2.0	1315.8	1307.9
ripe pulp	1311	2.7	1316.5	1306.1

Alpha diversity

In our analysis of average differences in compound richness across organs and species, we found a strong interaction between organ and species ($X^2 = 128.99$, $P < 0.0001$) and further examined differences among organs for each species separately. Organs often showed clear differences in average richness, but the patterns were highly variable across species (Fig. 4). In three species (*P. glabrescens, P. reticulatum,* and *P. slivivagum*), pulp and/or seeds had higher compound richness than leaves. However, in two species (*P. multiplinervum* and *P. generalense*) leaves had higher compound richness than all other fruit organs.

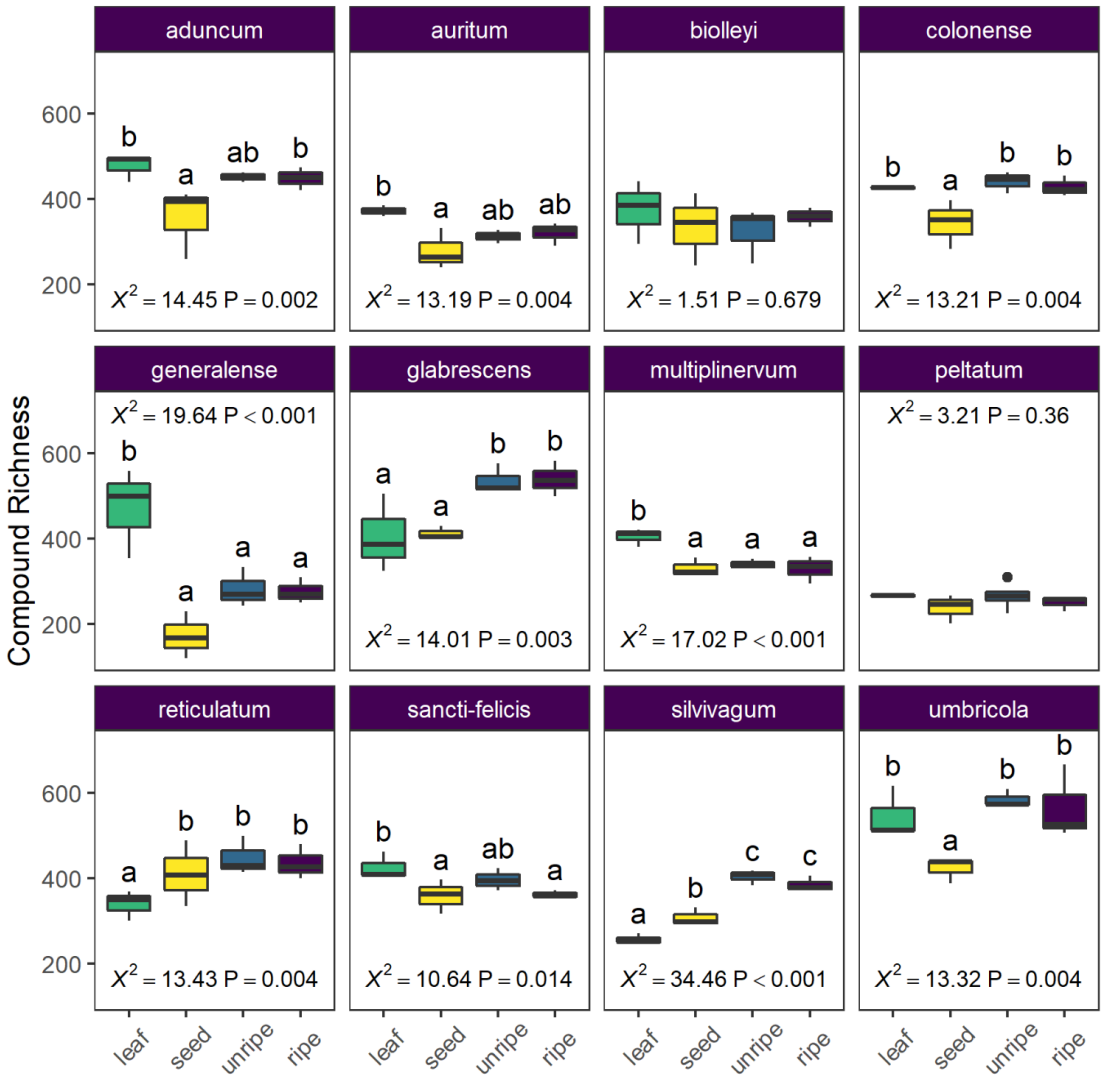
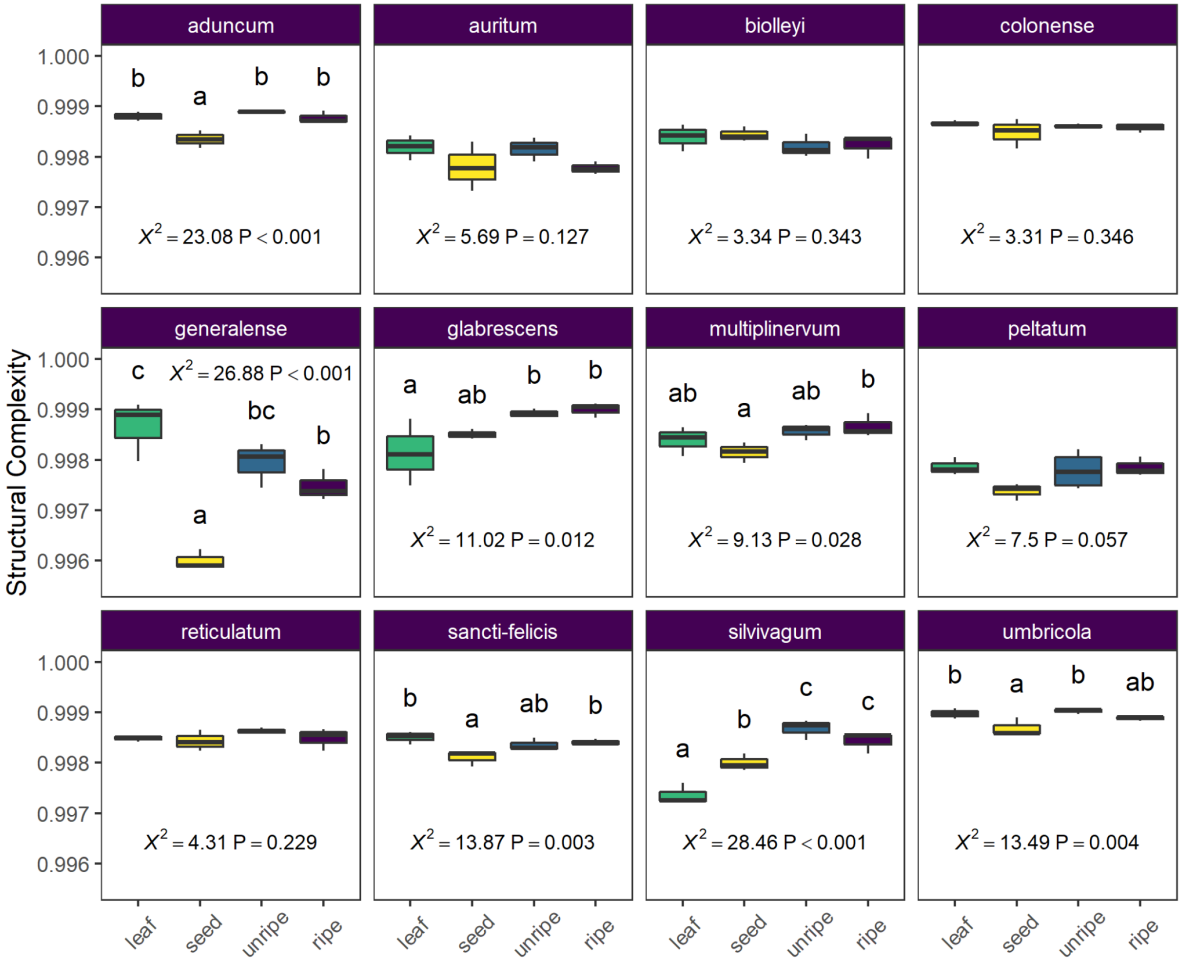


Figure 4: Average chemical richness differs across species and organ type (leaf, seed, unripe pulp, and ripe pulp). Letters indicate results of pairwise Tukey post-hoc comparisons of

493 organs within each species, with non-shared letters indicating a significant difference at $P <$
494 0.05. Each species plot includes X^2 and P-values from species-level LMMs.
495

496 ***Structural complexity***

497 In our analysis of average differences in structural complexity across organs and
498 species, we found a strong interaction between organ and species ($X^2 = 131.13$, $P < 0.0001$)
499 and further examined differences among organs for each species separately. For seven of
500 twelve species, organs showed differences in average structural complexity, but the patterns
501 were variable across species (Fig. 5). In two species (*P. glabrescens* and *P. slivivagum*), one
502 or more fruit organs had higher complexity than leaves. In another species (*P. generalense*),
503 leaves had higher complexity than all other fruit organs. Often, seeds had the lowest
504 structural complexity, or at least lower structural complexity than unripe or ripe fruit pulp
505 (Fig. 5).

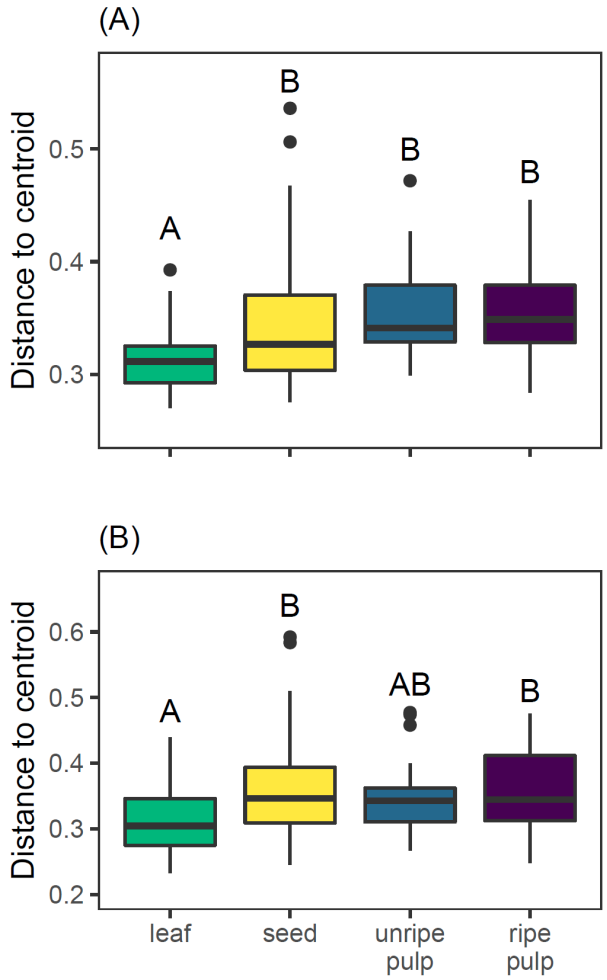


506

Figure 5: Average structural complexity differs across species and organ type (leaf, seed unripe pulp, and ripe pulp). Letters above each box plot column indicate results of pairwise Tukey post-hoc comparisons of organs within each species, with non-shared letters indicating a significant difference at $P < 0.05$. Each species plot includes X^2 and P-values from species-level LMMs.

Beta diversity

We found that beta-diversity in chemical composition was higher for fruits than leaves when considering only compound occurrences as well as structural composition. First, for compound occurrences, there was strong support for overall differences in beta diversity across organ types ($F_{3,139} = 7.56$, $P = 0.001$), with higher average distances to the group centroid for seeds, unripe pulp, and ripe pulp relative to leaves (Fig. 6A). Next, for structural composition, there was also strong support for overall differences in beta diversity across organ types ($F_{3,139} = 4.10$, $P = 0.009$). In this case, leaves had lower beta diversity than seeds or ripe pulp, and unripe pulp was intermediate (Fig. 6B). Further analyses conducted separately for each organ type showed that the differences in beta-diversity among organ types was due to variation both at the interspecific and intraspecific level (Table 4). A large proportion of the sample-to-sample variation within organ types (67-86%) was explained by differences among species relative to that explained by variation within species (14-33%), and this was especially true for unripe and ripe fruit pulp (Table 4).



528

529 **Figure 6:** Beta diversity in chemical composition is higher for reproductive organs than
530 leaves when considering variance in compound occurrences (A) or structural composition
531 (B). Letters indicate results of pairwise Tukey post-hoc comparisons of organs, with non-
532 shared letters indicating a significant difference at $P < 0.05$.

533

534

535

536

537

538

539

Table 4: Results from PERMANOVAs showing a large percentage of sample-to-sample variance in composition within organ types (i.e. beta diversity) is explained by species

	$F_{11,35}$[b]	P[b]	η^2 (Species)[c]	η^2 (Residual)[d]
Compound Occurrences [a]				
leaf	6.29	0.001	0.74	0.26
seed	5.17	0.001	0.70	0.30
unripe pulp	12.46	0.001	0.86	0.14
ripe pulp	13.18	0.001	0.86	0.14
Structural Composition [a]				
leaf	4.51	0.001	0.67	0.33
seed	6.58	0.001	0.75	0.25
unripe pulp	7.90	0.001	0.79	0.21
ripe pulp	10.70	0.001	0.83	0.17

[a] Separate sets of PERMANOVAs were conducted for each aspect of compound composition: compound occurrences (presence/absence) and structural composition

[b] Statistical results from permutation tests showing strong support for an effect of Piper species on composition

[c] Proportion of sample-to-sample variance explained by species (i.e. interspecific variation)

[d] Proportion of sample-to-sample variance explained by individual and within-individual (residual) variance

Discussion

Across the many angiosperms with animal-dispersed seeds, the functional traits of fruits have been shaped by particularly complex selective pressures, imposed in part by antagonistic and mutualistic consumers. Plant secondary metabolites are prominent among the traits that have evolved to mediate plant-consumer interactions. As such, it can be anticipated that the secondary metabolites of fruits in animal-dispersed plants will reflect the many facets of their ecological and evolutionary settings -- in their diversity of chemical composition, molecular structures, and ecological functions. In this study, we surveyed and compared the compositional and structural diversity of secondary metabolites across vegetative and reproductive organs in 12 species of the genus *Piper*. In all metrics of secondary metabolite diversity that were quantified, the overall diversity of fruit organs matched or exceeded those of leaves, though these patterns varied across species. The patterns of secondary metabolite diversity that were revealed across vegetative and reproductive organs are in line with the multifarious functional roles that have been hypothesized for secondary metabolites in fruits.

Our untargeted metabolomic survey of phytochemical occurrence patterns revealed that fruit organs harbor fruit-specific metabolites from a variety of chemical classes (Table 1, Fig. 1), in each class equal to or greater in number than those that were leaf-specific (Table 1). This included classes of compounds that have previously been found to be more numerous and abundant in fruit organs (e.g. amides; Whitehead *et al.*, 2013), as well as numerous chemical classes previously described in studies of *Piper* spp. leaf chemistry (Parmar *et al.*, 1997; Baldoqui *et al.*, 1999; Kato & Furlan, 2007; Richards *et al.*, 2015). The occurrence of numerous fruit-specific secondary metabolites from a variety of unlinked biosynthetic pathways suggests a pattern of fruit-specific secondary metabolite trait evolution, likely a result of fruit-specific selective pressures.

The evolution of organ-specific phytochemical traits across our target plant species is also made evident by the results of our machine learning analysis. Here, our random forest model was very successful at distinguishing among organ types based solely on their secondary metabolite composition. Most notably, the exceptional performance of the classification algorithm to distinguish between vegetative and reproductive organs can only be explained by the presence of strong association between chemical composition and organ type. Despite the fact that our species set included vines, understory shrubs, and pioneering taxa, all adapted to very different local habitats, these associations are consistent across all 12 focal species.

The clustering patterns found by our NMDS analysis (Fig. 2) show clustering at two different levels. First, the samples from different organs from the same species cluster together. This pattern strongly suggests the presence of physiological or genetic linkage constraints in the organ specific evolution of phytochemicals. The strong chemical similarity across organs within a species could point toward the influence that changes in the expression or composition of secondary compounds in one plant organ could have on the expression or composition of other organs. Although our data do not allow us to disentangle the precise mechanisms that give rise to these patterns, it is clear that chemical changes in one plant organ are likely to be mirrored, to some extent, by changes in the chemical architecture of the whole plant. Second, as expected, and despite the strong chemical similarity exhibited by organs within a *Piper* species, samples also show a clear pattern or clustering by organ type (Fig. 2). This pattern reinforces the expectation that the distinctive regimes of selective pressures imposed upon the different plant organs are sufficiently strong to create convergent organ-specific patterns of the chemical composition, and that these selective regimes are likely to be consistent across species and habitats. The Boruta variable importance model, a widely used machine learning algorithm designed to identify statistically important classification variables from large datasets, revealed specific compounds from at least six different chemical classes as key features that distinguish vegetative and reproductive organs (Table S1).

While the majority of significant Boruta variables, like the overall majority of secondary metabolites cataloged in our study, exhibited some overlap in occurrence across leaf and fruit organs when the 12 *Piper* species were evaluated as a group (Fig. 3A), there was substantially less overlap at the level of individual species (Fig. S1). In many cases, these patterns of variance were the result of numerous compounds occurring in only one organ type in a certain species or subset of species, but occurring more widely in another species or subset of species.

The broad overlap across organs in compound occurrence at the genus level provides a degree of insight into the extent of constraints on organ-specific chemical trait evolution at this taxonomic scale. However, to a degree this overlap can also be attributed to the shared demand for defensive compounds across vegetative and reproductive organ types. While phylogenetic data will be required in order to infer the ancestral organ localizations of phytochemical traits of *Piper*, the widespread variation in organ localizations that we observed across species suggests that genetic constraints have not bound these traits to a

certain organ type over the course of *Piper* speciation. Further, the apparent mobility of secondary metabolite traits across organ types within the genus suggests a bidirectional exchange of these traits, which, when vegetative and reproductive organs are each threatened by separate assemblages of consumers, may allow more rapid defense trait adaptation than can arise from novel mutations.

Our untargeted metabolomic survey has shown that fruit organs are at the very least a reservoir of phytochemical richness. While the alpha diversity of organs at the species level was highly variable (Fig. 4), rarefaction analysis of gamma diversity showed a small but clear trend towards higher richness of secondary metabolites in reproductive organs (Fig. 3; Table 3). Similarly, while chemical structural complexity of organs at the species level was highly variable (Fig. 5), chemical structural variance (β-diversity) across species was significantly higher for reproductive organs than for leaves (Fig. 6). In summary, these trends indicate not only that reproductive organs accumulate a higher number of secondary metabolite traits than do leaves, but also that these traits are more divergent from one another across species than those of leaves. These trends are consistent with higher overall evolutionary diversification of phytochemical traits in reproductive organs, suggesting that fruits may be an important, but underappreciated, force in shaping chemical trait evolution at the whole plant level.

Acknowledgements

This research was supported by National Science Foundation (Grants No. DEB-1210884 and DEB-1856776 to SRW) and start-up funds to SRW from the Virginia Tech Department of Biological Sciences. The mass spectrometry resources used in this work were maintained with funds from the Fralin Life Science Institute as well as the Virginia Agricultural Experiment Station Hatch Program (VA-160085). We thank the Organization for Tropical Studies for logistical support at La Selva Biological Station. Marisol Luna Martinez assisted with sample collection and Orlando Vargas Ramírez assisted with plant identification. Natalie Rodeman and Katherine Berg assisted with sample processing for chemical analyses.

Author contributions

GFS, SRW, and DS designed the research; SRW collected field samples; GFS, SBH, and RFH conducted chemical analyses, GFS conducted molecular networking and data curation; GFS and SRW conducted the statistical analysis with contributions from DS; GFS

wrote the first draft of the manuscript with contributions from SRW, and all authors contributed substantially to revisions and approved the final version.

Data availability

All data and R scripts will be available through Dryad digital repository upon publication.

References

Adler LS, Seifert MG, Wink M, Morse GE, Turlings T. **2012**. Reliance on pollinators predicts defensive chemistry across tobacco species. *Ecology Letters* **15**: 1140–1148.

Adler LS, Wink M, Distl M, Lentz AJ. **2006**. Leaf herbivory and nutrients increase nectar alkaloids. *Ecology Letters* **9**: 960–7.

Aron AT, Gentry EC, McPhail KL, Nothias LF, Nothias-Esposito M, Bouslimani A, Petras D, Gauglitz JM, Sikora N, Vargas F, *et al.* 2020. Reproducible molecular networking of untargeted mass spectrometry data using GNPS. *Nature Protocols* **15**: 1954–1991.

Benton HP, Want EJ, Ebbels TMD. 2010. Correction of mass calibration gaps in liquid chromatography-mass spectrometry metabolomics data. *BIOINFORMATICS* **26**: 2488.

Baldoqui DC, Kato MJ, Cavalheiro AJ, Bolzani VDS, Young MCM, Furlan M. 1999. A chromene and prenylated benzoic acid from *Piper aduncum*. *Phytochemistry* **51**: 899–902.

Berardi AE, Hildreth SB, Helm RF, Winkel BSJ, Smith SD. **2016**. Evolutionary correlations in flavonoid production across flowers and leaves in the Iochrominae (Solanaceae). *Phytochemistry* **130**: 119–127.

Brehm G, Hebert PDN, Colwell RK Adams MO, Bodner F, Friedemann K, Möckel L,Fiedler K. 2016. Turning up the heat at a hotspot: DNA barcodes reveal 80% more species of geometrid moths along an Andean elevational gradient. *PlosOne* **11:** e0150327. doi: 10.1371/journal.pone.0150327

671 **Breiman L. 2001.** Random forests. *Machine learning* **45**: 5-32.

672 **Breiman L. 2002.** Manual on setting up, using, and understanding random forests v3. 1.
673 URL: www.stat.berkeley.edu/~breiman/Using_random_forests_V3.1.pdf [accessed 20
674 January 2021].

675 **Cech NB, Enke CG. 2001.** Practical implications of some recent studies in electrospray
676 ionization fundamentals. *Mass Spectrometry Reviews* **20**: 362–387.

677 **Chen H, Köllner TG, Li G, Wei G, Chen X, Zeng D, Qian Q, Chen F. 2020.**
678 Combinatorial evolution of a terpene synthase gene cluster explains terpene variations in
679 *Oryza*. *Plant Physiology* **182**: 480–492.

680 **Chiarucci A, Bacaro G, Rocchini D, Ricotta C, Palmer M, Scheiner S. 2009.** Spatially
681 constrained rarefaction: Incorporating the autocorrelated structure of biological
682 communities into sample-based rarefaction. *Community Ecology* **10**: 209–214.

683 **Christian N, Sedio BE, Florez-Buitrago X, Ramírez-Camejo LA, Rojas EI, Mejía LC,**
684 **Palmedo S, Rose A, Schroeder JW, Herre EA. 2020.** Host affinity of endophytic fungi
685 and the potential for reciprocal interactions involving host secondary chemistry. *American*
686 *Journal of Botany* **107**: 219–228.

687 **Dyer LA, Palmer ADN**. **2004**. *Piper: A model Genus for Studies of Phytochemistry,*
688 *Ecology, and Evolution*. Kluwer Academic/ Plenum Publishers.

689 **Dyer LA, Philbin CS, Ochsenrider KM, Richards LA, Massad TJ, Smilanich AM,**
690 **Forister ML, Parchman TL, Galland LM, Hurtado PJ*, et al.* 2018**. Modern approaches
691 to study plant - insect interactions in chemical ecology. *Nature Reviews Chemistry* **2**: 50–
692 64.

693 **Falara V, Akhtar TA, Nguyen TTH, Spyropoulou EA, Bleeker PM, Schauvinhold I,**
694 **Matsuba Y, Bonini ME, Schilmiller AL, Last RL*, et al.* 2011.** The tomato terpene
695 synthase gene family. *Plant Physiology* **157**: 770–789.

696 **Gentry AH. 1990.** *Four neotropical rainforests.* Yale University Press, New Haven.

697 **Gotelli NJ, Colwell RK. 2011.** Estimating species richness. In: Magurran AE, McGill BJ,
698 eds. *Biological Diversity: Frontiers in Measurement and Assessment*. Oxford, UK: Oxford
699 University Press, 39–54.

700 **Greig N. 1993.** Predispersal seed predation on five *Piper* species in tropical rain-forest.
701 *Oecologia* **93**: 412–420

702 **Hartmann T**. **2007**. From waste products to ecochemicals: Fifty years research of plant
703 secondary metabolism. *Phytochemistry* **68**: 2831–2846.

704 **Herrera CM. 1982.** Defense of ripe fruit from pests: Its significance in relation to plant-
705 disperser interactions. *The American Naturalist* **120**: 218–241.

706 **Iason GR, Dicke M, Hartley SE. 2012.** *The ecology of plant secondary metabolites: From*
707 *genes to global processes.* Cambridge, UK: Cambridge University Press.

708 **Kato MJ, Furlan M**. **2007**. Chemistry and evolution of the Piperaceae. *Pure and Applied*
709 *Chemistry* **79**: 529–538.

710 **Keith RA, Mitchell-Olds T**. **2019**. Antagonistic selection and pleiotropy constrain the
711 evolution of plant chemical defenses. *Evolution* **73**: 947–960.

712 **Kessler A, Baldwin IT**. **2002**. Plant responses to insect herbivory: The emerging molecular
713 analysis. *Annual Review of Plant Biology* **53**: 299–328.

714 **Kessler A, Halitschke R**. **2009**. Testing the potential for conflicting selection on floral
715 chemical traits by pollinators and herbivores: Predictions and case study. *Functional*
716 *Ecology* **23**: 901–912.

717 **Kuhl C, Tautenhahn R, Boettcher C, Larson TR, Neumann S. 2012.** CAMERA: an
718 integrated strategy for compound spectra extraction and annotation of liquid
719 chromatography/mass spectrometry data sets. *Analytical Chemistry*, **84**, 283–289.

720 **Kursa MB, Rudnicki WR. 2010.** Feature selection with the Boruta package. *Journal of*
721 *Statistical Software* **36**: 1-13.
722 **Kursar TA, Dexter KG, Lokvam J, Pennington RT, Richardson JE, Weber MG,**
723 **Murakami ET, Drake C, McGregor R, Coley PD. 2009**. The evolution of antiherbivore
724 defenses and their contribution to species coexistence in the tropical tree genus *Inga*.
725 *Proceedings of the National Academy of Sciences of the United States of America* **106**:
726 18073–18078.
727 **Liaw A, Wiener M. 2002.** Classification and regression by randomForest. *R News* **2**: 18-22.
728 **Martinez Arbizu P. 2020**. pairwiseAdonis: Pairwise multilevel comparison using adonis. R
729 package version 0.4. URL https://github.com/pmartinezarbizu/pairwiseAdonis . [accessed
730 9 January 2021].
731 **Maynard LD, Slinn HL, Glassmire AE, Matarrita-Carranza B, Dodson CD, Nguyen**
732 **TT, Burroughs MJ, Dyer LA, Jeffrey CS, Whitehead SR. 2020.** Secondary metabolites
733 in a neotropical shrub: spatiotemporal allocation and role in fruit defense and dispersal.
734 *Ecology* **101**: 1–15.
735 **Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR,**
736 **O'Hara RB, Simpson GL, Solymos P, *et al.* 2019.** vegan: Community Ecology Package.
737 R package version 2.5-6. URL https://CRAN.R-project.org/package=vegan . [accessed 9
738 January 2021]
739 **Parmar VS, Jain SC, Bisht KS, Jain R, Taneja P, Jha A, Tyagi OD, Prasad AK, Wengel**
740 **J, Olsen CE, et al. 1997.** Phytochemistry of the genus *Piper*. *Phytochemistry* **46**: 597–
741 673.
742 **Richards LA, Dyer LA, Forister ML, Smilanich AM, Dodson CD, Leonard MD, Jeffrey**
743 **CS. 2015**. Phytochemical diversity drives plant–insect community diversity. *Proceedings*
744 *of the National Academy of Sciences of the United States of America* **112**: 10973–10978.

745 **Ricigliano VA, Sica VP, Knowles SL, Diette N, Howarth DG, Oberlies NH. 2020.**
746 Bioactive diterpenoid metabolism and cytotoxic activities of genetically transformed
747 Euphorbia lathyris roots. *Phytochemistry*. doi:10.1016/j.phytochem.2020.112504
748 **Salazar D, Jaramillo MA, Marquis RJ**. **2016**. Chemical similarity and local community
749 assembly in the species rich tropical genus *Piper*. *Ecology* **97**: 3176–3183.
750 **Salazar D, Lokvam J, Mesones I, Pilco MV, Milagros J, Zuñiga A, Valpine P De, Fine**
751 **PVA**. **2018**. Origin and maintenance of chemical diversity in a species-rich tropical tree
752 lineage. *Nature Ecology & Evolution* **2**: 983–990.
753 **Schneider GF, Coley PD, Younkin GC, Forrister DL, Mills AG, Kursar TA. 2019**.
754 Phenolics lie at the centre of functional versatility in the responses of two phytochemically
755 diverse tropical trees to canopy thinning. *Journal of Experimental Botany* **70**: 5853–5864.
756 **Sedio BE, Echeverri JCR, Boya CA, Wright SJ**. **2017**. Sources of variation in foliar
757 secondary chemistry in a tropical forest tree community. *Ecology* **98**: 616–623.
758 **Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. 2006.** XCMS: Processing mass
759 spectrometry data for metabolite profiling using nonlinear peak alignment, matching and
760 identification. *Analytical Chemistry* **78**: 779–787.
761 **Suzuki T, Iwai K. 1984.** Constituents of red pepper species: Chemistry, biochemistry,
762 pharmacology, and food science of the pungent principle of Capsicum species. In: Brossi
763 A, ed. *The Alkaloids, Volume 23.* Academic Press, 227–299.
764 **Tautenhahn R, Boettcher C, Neumann S. 2008.** Highly sensitive feature detection for high
765 resolution LC/MS. *BMC Bioinformatics*. doi: 10.1186/1471-2105-9-504.
766 **Tewksbury JJ. 2002.** Fruits, frugivores and the evolutionary arms race. *New Phytologist*
767 **156**: 137–139.
768 **Tewksbury JJ, Nabhan GP**. **2001**. Seed dispersal: Directed deterrence by capsaicin in
769 chillies. *Nature* **412**: 403–404.

770 **Tewksbury JJ, Reagan KM, Machnicki NJ, Carlo TA, Haak DC, Peñaloza ALC, Levey**
771 **DJ**. **2008**. Evolutionary ecology of pungency in wild chilies. *Proceedings of the National*
772 *Academy of Sciences of the United States of America* **105**: 11808–11811.
773 **Thies W, Kalko EKV. 2004.** Phenology of Neotropical pepper plants (Piperaceae) and their
774 association with their main dispersers, two short-tailed fruit bats, *Carollia perspicillata*
775 and *C. castanea* (Phyllostomidae). *Oikos* **104**: 362–376.
776 **Volf M, Segar ST, Miller SE, Isua B, Sisol M, Aubona G, Šimek P, Moos M, Laitila J,**
777 **Kim J, *et al.* 2018**. Community structure of insect herbivores is driven by conservatism,
778 escalation and divergence of defensive traits in *Ficus*. *Ecology Letters* **21**: 83–92.
779 **Wang M, Carver JJ, Phelan V V, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous**
780 **J, Kapono CA, Luzzatto-Knaan T, *et al.* 2016**. Sharing and community curation of mass
781 spectrometry data with Global Natural Products Social Molecular Networking. *Nature*
782 *Biotechnology* **34**: 828–837.
783 **Wetzel WC, Whitehead SR**. **2020**. The many dimensions of phytochemical diversity:
784 linking theory to practice. *Ecology Letters* **23**: 16–32.
785 **Whitehead SR, Bowers MD. 2013a.** Iridoid and secoiridoid glycosides in a hybrid complex
786 of bush honeysuckles (*Lonicera* spp., Caprifolicaceae): Implications for evolutionary
787 ecology and invasion biology. *Phytochemistry* **86**:57–63.
788 **Whitehead SR, Bowers MD. 2013b.** Evidence for the adaptive significance of secondary
789 compounds in vertebrate-dispersed fruits. *The American Naturalist* **182**: 563–577.
790 **Whitehead SR, Bowers MD. 2014.** Chemical ecology of fruit defence: synergistic and
791 antagonistic interactions among amides from *Piper*. *Functional Ecology* **28**: 1094–1106.
792 **Whitehead SR, Jeffrey CS, Leonard MD, Dodson CD, Dyer LA, Bowers MD. 2013.**
793 Patterns of secondary metabolite allocation to fruits and seeds in *Piper reticulatum*.
794 *Journal of Chemical Ecology* **39**: 1373–1384.

795 **Whitehead SR, Quesada MFO, Bowers MD**. **2016**. Chemical tradeoffs in seed dispersal:
796 defensive metabolites in fruits deter consumption by mutualist bats. *Oikos* **125**: 927–937.
797 **Wittstock U, Gershenzon J**. **2002**. Constitutive plant toxins and their role in defense against
798 herbivores and pathogens. *Current Opinion in Plant Biology* **5**: 300–307.
799 **Xu S, Kreitzer C, McGale E, Lackus ND, Guo H, Köllner TG, Schuman MC, Baldwin**
800 **IT, Zhou W. 2020.** Allelic differences of clustered terpene synthases contribute to
801 correlated intraspecific variation of floral and herbivory-induced volatiles in a wild
802 tobacco. *New Phytologist*. doi: 10.1111/nph.16739